NOTE Communicated by Garrison Cottrell and Alice O'Toole

A Perceptron Reveals the Face of Sex

Michael S. Gray
David T. Lawrence
Beatrice A. Golomb
Terrence J. Sejnowski
Howard Hughes Medical Institute,
Computational Neurobiology Laboratory, The Salk Institute for Biological Studies,
P.O. Box 85800, San Diego, CA 92186-5800 USA and
Departments of Biology and Cognitive Science,
University of California, San Diego, La Jolla, CA 92093 USA

Recognizing the sex of conspecifics is important. Humans rely primarily on visual pattern recognition for this task. A wide variety of linear and nonlinear models have been developed to understand this task of sex recognition from human faces.[1] These models have used both pixel-based and feature-based representations of the face for input. Fleming and Cottrell (1990) and Golomb *et al.* (1991) utilized first an autoencoder compression network on a pixel-based representation, and then a classification network. Brunelli and Poggio (1993) used a type of radial basis function network with geometrical face measurements as input. O'Toole and colleagues (1991, 1993) represented faces as principal components. When the hidden units of an autoencoder have a linear output function, then the N hidden units in the network span the first N principal components of the input (Baldi and Hornik 1989). Bruce *et al.* (1993) constructed a discriminant function for sex with 2-D and 3-D facial measures.

In this note we compare the performance of a simple perceptron and a standard multilayer perceptron (MLP) on the sex classification task. We used a range of spatial resolutions of the face to determine how the reliability of sex discrimination is related to resolution. A normalized pixel-based representation was used for the faces because it explicitly retained texture and shape information while also maintaining geometric relationships. We found that the linear perceptron model can classify sex from facial images with 81% accuracy, compared to 92% accuracy with compression coding on the same data set (Golomb *et al.* 1991). The advantage of using a simple linear perceptron with normalized pixel-based inputs is that it allows us to see explicitly those regions of the face

[1] Consistent with Burton *et al.* (1993), we use the term sex rather than gender because our interest is in the physical, not psychological, characteristics of the face.

that make the largest and most reliable contributions to the classification of sex.

A database of 90 faces (44 males, 46 females) was used (O'Toole *et al.* 1988). No facial hair, jewelry, or makeup was on any of the faces. Each face was rotated until the eyes were level, and then scaled and cropped so that each image showed a similar facial area. From the original set of faces, we created five separate databases at five different resolutions (10 × 10 pixels, 15 × 15, 22 × 22, 30 × 30, and 60 × 60). To produce each of these databases, the original faces were deterministically subsampled by selecting pixels from the original image at regular intervals. For all faces at a given resolution, the images were equalized for brightness from the initial 256 gray-levels. A sample face is shown in Figure 1a. Because not all photos were exactly head-on, each database was doubled in size (to 180 faces) by flipping each face image horizontally and including this new image in the database. This procedure removed any systematic lateral differences would could have been exploited by the network.

Two different architectures were used: (1) a simple perceptron and (2) an MLP. In the simple perceptron model, the inputs (the face image) were directly connected to a single output unit. The MLP model included a layer of 10 hidden units between the input and output units.

A jackknife training procedure was used. For each architecture at each resolution, 9 separate networks were trained. Each of these 9 networks was trained on a different subset containing 160 of the 180 faces, with the remaining 20 test faces used to measure generalization performance. These 20 test faces constituted a unique testing set for each network, and consisted of 10 individuals with their horizontally flipped mirror images. There was, of course, a high degree of overlap in the faces used for training the different networks. The networks were trained with conjugate gradient until all patterns in the training set were within 0.2 of the desired output (1.0 for males, 0.0 for females) activation, or until network performance was not improving.

The simple perceptron and the MLP demonstrated remarkably similar generalization performance at all resolutions (see Fig. 1b). Comparison of the performance of the two architectures within each resolution revealed no significant differences ($p > 0.05$ in all cases). There was, however, a significant improvement at higher resolution for the perceptron networks [$F(4,40) = 3.121$, $p < 0.05$] and for the MLP networks [$F(4,40) = 3.789$, $p < 0.05$]. Post-hoc comparisons showed that for the perceptron networks, generalization performance at a resolution of 10 × 10 pixels was significantly worse than at all other (higher) resolutions. For the MLP networks, performance also degraded with lower resolutions. The 10 × 10 MLP networks were significantly worse than the 22 × 22, 30 × 30, and 60 × 60 networks; the 15 × 15 networks were significantly worse than the 30 × 30 networks.

Examination of the weights of the perceptron network revealed how the solution was reached. Figure 1c shows the mean weights of the

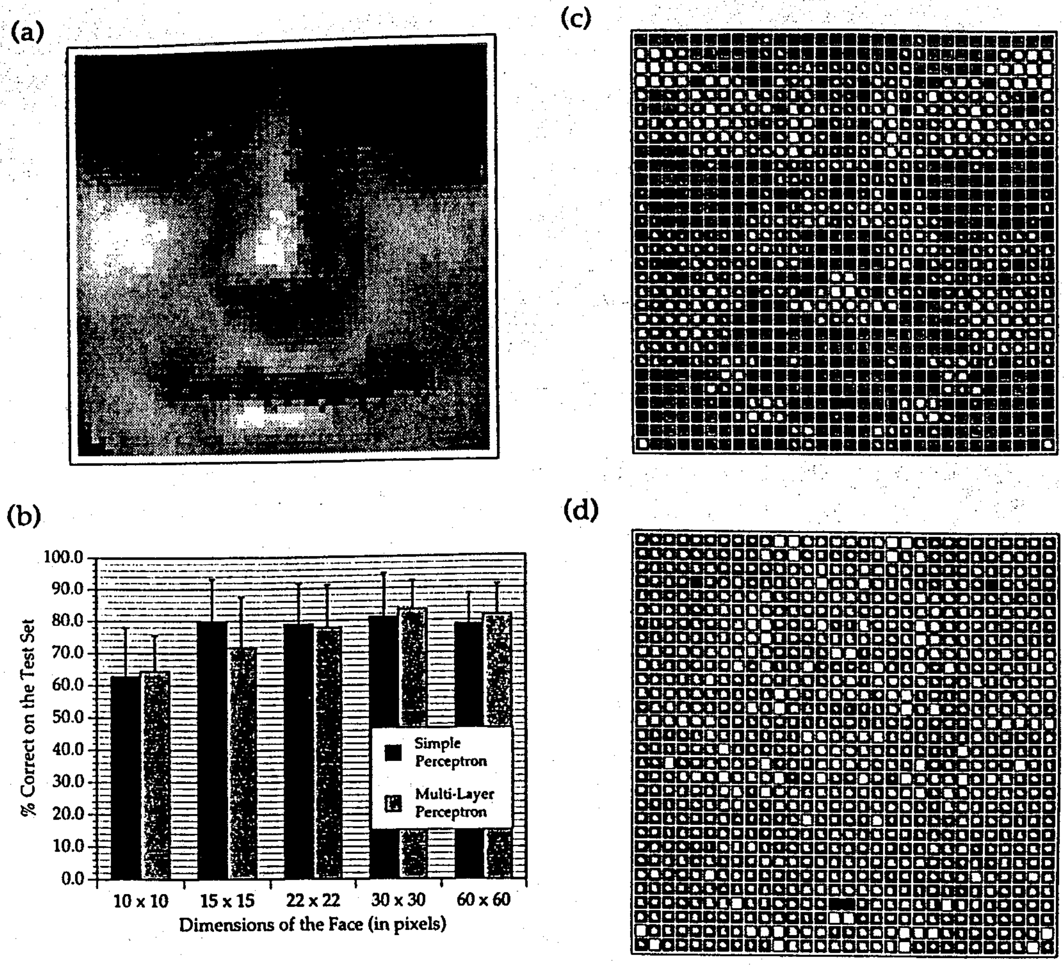

Figure 1: (a) Sample face from database; (b) performance of two types of networks and different input sizes; (c) weights in a 30 × 30 perceptron network; (d) logarithm of the coefficient of variation of the weight in (c).

9 simple perceptron networks (30 × 30 pixel resolution) at the end of training. Figure 1d shows $\log(\sigma_w/|w|)$, the logarithm of the coefficient of variation (the standard deviation of the weight divided by its absolute mean value).

Recent efforts to match human performance on sex recognition have been remarkably successful. Using the same network architecture but with different training sets, Fleming and Cottrell (1990) had an accuracy rate of 67%, while Golomb *et al.* (1991) achieved model generalization performance of 91.9% correct, compared to 88.4% for humans. Using a leave-one-out training strategy, Brunelli and Poggio (1993) demonstrated 87.5% correct generalization performance. Burton *et al.* (1993) constructed a discriminant function using a variety of 2-D and 3-D face measurements. They achieved 85.5% accuracy over their set of 179 faces using 12 simple measurements from full-face (frontal) photographs. With 16 2-D and 3-D variables, their performance improved to 93.9%. It is important to note, however, that this is not a generalization measure for new faces, but indicates training performance on their complete set of faces. O'Toole *et al.* (1991) reached generalization performance of 74.3% accuracy when combining information from the first four eigenvectors.

Compared to these previous studies, our performance of 81% with the simple perceptron is not exceptional. There are, however, several important aspects to our approach. First, we use a normalized pixel-based input. With the normalization, we bring the eyes and mouth into exact register, and limit the range of luminance values in the image. Another advantage of our pixel-based approach (as opposed to geometric measurements) is that all regions of the face are represented in the input. Through training, the network determines which parts of the face have reliable information, and which are less consistent. When one chooses to represent faces as (arbitrary) geometric measurements, however, information is lost from the beginning. Regardless of the model used subsequently to classify the faces, it can use only the measurements collected. Our method does not depend on intuition regarding which regions or features of the face are important.

The particular advantage of the perceptron model is that it shows explicitly how the sex classification problem was solved. Figure 1c shows that the nose width and image intensity in the eye region are important for males while image intensity in the mouth and nose area is important for discriminating women. In Figure 1d, showing the logarithm of the coefficient of variation of the weights across networks, most regions seem to provide reliable information (small squares). There are a few areas (e.g., the outside of the nose) that have particularly high variability (large squares) across networks. More importantly for both of figure 1c and 1d, however, we see that information relevant to sex classification is broadly distributed across all regions of the face.

Our results show that a simple perceptron architecture was found to perform as well as an MLP on a sex classification task with normalized pixel-based inputs. Performance was also surprisingly good even at the coarser resolutions tested. The high degree of similarity between the results of the two architectures suggests that a substantial part of the problem is linearly separable, consistent with the results of O'Toole and colleagues (1991, 1993) using principal components. This simple perceptron, with less than 2% of the number of parameters in the model by Golomb *et al.* (1991), reached a peak performance level of 81% correct. Since human performance on the same faces is around 88%, sex recognition may in fact be a simpler skill than previously believed.

Acknowledgments

We thank David C. Rogers and Marion Stewart Bartlett for technical assistance, and two reviewers for their helpful comments. This study was supported by grants from the National Science Foundation and the Office of Naval Research.

References

Baldi, P., and Hornik, K. 1989. Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks* **2**, 53–58.

Bruce, V., Burton, M. A., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., and Linney, A. 1993. Sex discrimination: How do we tell the difference between male and female faces? *Perception* **22**, 131–152.

Brunelli, R., and Poggio, T. 1993. Caricatural effects in automated face perception. *Biol. Cybernet.* **69**, 235–241.

Burton, M. A., Bruce, V., and Dench, N. 1993. What's the difference between men and women? Evidence from facial measurement. *Perception* **22**, 153–176.

Fleming, M., and Cottrell, G. W. 1990. Categorization of faces using unsupervised feature extraction. In *Proceedings of IJCNN-90*, Vol. 2, pp. 65–70. IEEE Neural Networks Council, Ann Arbor, MI.

Golomb, B. A., Lawrence, D. T., and Sejnowski, T. J. 1991. Sexnet: A neural network identifies sex from human faces. In *Advances in Neural Information Processing Systems*, R. P. Lippman, J. Moody, and D. S. Touretzky, eds., Vol. 3, pp. 572–577. Morgan Kaufmann, San Mateo, CA.

O'Toole, A. J., Millward, R. B., and Anderson, J. A. 1988. A physical system approach to recognition memory for spatially transformed faces. *Neural Networks* **1**, 179–199.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A., and Bartlett, J. C. 1991. Classifying faces by race and sex using an autoassociative memory trained for recognition. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, K. J. Hammon and D. Getner, eds., Vol. 13, pp. 847–885. Lawrence Erlbaum, Hillsdale, NJ.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A., and Valentin, D. 1993. Low-dimensional representation of faces in higher dimensions of the face space. *J. Opt. Soc. Am. A* **10**(3), 405–411.

Received January 6, 1994; accepted January 30, 1995.